GCNs-Based Context-Aware Short Text Similarity Model

Xiaoqi Sun, Shaochun Wu, Yue Liu

Shanghai University

Abstract

Semantic textual similarity is a fundamental task in text mining and natural language processing (NLP), which has profound research value. The essential step for text similarity is text representation learning. Recently, researches have explored the graph convolutional network (GCN) techniques on text representation, since GCN does well in handling complex structures and preserving syntactic information. However, current GCN models are usually limited to very shallow layers due to the vanishing gradient problem, which can not capture non-local dependency information of sentences. In this paper, we propose a GCNs-based context-aware (GCSTS) model that applies iterated GCN blocks to train deeper GCNs. Recurrently employing the same GCN block prevents over-fitting and provides broad effective input width. Combined with dense connections, GCSTS can be trained more deeply. Besides, we use dynamic graph structures in the block, which further extends the receptive field of each vertex in graphs, learning better sentence representations. Experiments show that our model outperforms existing models on several text similarity datasets, while also verify that GCNs-based text representation models can be trained in a deeper manner, rather than being trained in two or three layers.

Introduction

- Recently, Graph Convolutional Network (GCN) has attracted extensive attention. GCNs can extract semantic and syntactic information of sentences simultaneously from sentence dependency trees.
- CNs capture information only about immediate neighbors with one layer of convolution. L layers will be needed in order to capture neighborhood information that is L hops away. Therefore, a shallow GCN model can not be able to capture non-local interactions of long sentences. Notably, while deeper GCNs can capture more abundant neighborhood information of a graph, current GCN models used in NLP are no more than three layers. Due to the fact that the deeper GCN model introduces a higher complexity in backpropagation, and the vanishing gradients pose limitations on the depth growth of GCNs-based networks.
- Furthermore, most GCNs employ fixed graph structures. However, dynamic graph convolution, where the graph structure is allowed to change in each layer, can extend the receptive field of each vertex in the graph and learn a better graph representation compared to GCNs with fixed graph structure.
- On the basis of analysis above, we propose a deeper GCNs-based context-aware network to get better text representations. Our overall iterated GCN architecture repeatedly applies the same block of graph convolution to vertex-wise representations. Besides, we use dynamic graph structures in the GCN block. The dynamic graph convolution makes the representations of nodes influenced by non-fixed neighborhoods, which means the representations can bring more contextual information.
- In order to alleviate the vanishing gradient problem in deeper GCNs, we adapt a similar idea of DenseNet to our GCSTS model. DenseNet provided a big step forward in the pursuit of deep models when it introduced dense connections among layers. These connections massively alleviate the vanishing gradient problem. With the help of dense connections, we are able to train a deeper model, allowing rich local and non-local information to be captured.

Measures

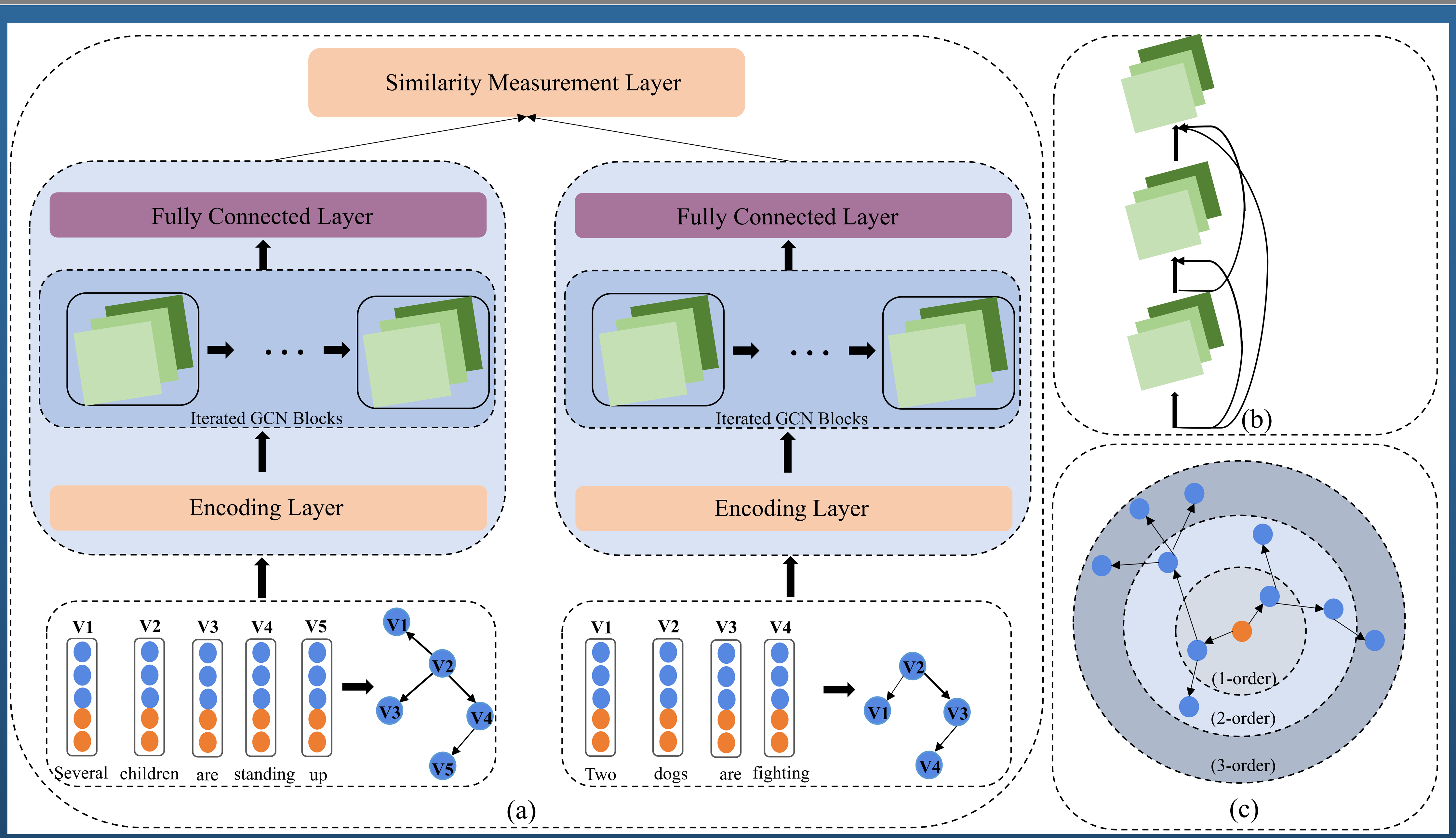

Figure 1. (a): The overall architecture of GCSTS with an example sentence pair and the dependency trees. GCSTS uses the siamese framework, which means the left and right network share weights. The network takes node embeddings and adjacency matrices that represent the graph as the input. GCSTS includes encoding layer, iterated GCN block, fully connected layer and similarity layer. (b): The iterated GCN blocks with dense connections. Each iteration takes the concatenation of the initial contextual representation and the node representations produced in proceeding iterations as input. (c): An example node and its 1-order, 2-order, and 3-order neighbors.

Results

Table 1
STS12-STS16: SemEval 2012-2016 datasets. STS-B: STSbenchmark dataset. We report the Spearman correlation coefficient in this work. The best results are bold.

Model	STS12	STS13	STS14	STS15	STS16	STS-B	Avg.
Avg. GloVe embeddings (Pennington et al., 2014))	0.547	0.711	0.585	0.653	0.630	0.575	0.617
Multi-Perspective-CNN (He et al., 2015)	0.571	0.725	0.627	0.702	0.666	0.594	0.647
Dependency Tree-LSTM (Tai et al., 2015)	0.583	0.759	0.649	0.731	0.685	0.616	0.671
Constituent Tree-LSTM (Tai et al., 2015)	0.577	0.746	0.645	0.727	0.673	0.607	0.663
Siamese LSTM (Mueller et al., 2016)	0.580	0.758	0.644	0.724	0.689	0.610	0.668
InferSent (Conneau et al., 2017)	0.599	0.763	0.650	**0.748**	0.694	0.625	0.679
Avg. BERT embeddings (Devlin et al., 2019)	0.520	0.693	0.584	0.650	0.615	0.562	0.604
Text-GNN (Huang et al., 2019)	0.603	0.771	0.656	0.736	0.710	0.624	0.683
GCSTS	**0.615**	**0.782**	**0.669**	0.742	**0.715**	**0.637**	**0.693**

Table 2 Experimental results on the MRPC dataset. The best results are bold.

Model	Accuracy	F1
Avg. GloVe embeddings (Pennington et al., 2014))	0.721	0.749
Multi-Perspective-CNN (He et al., 2015)	0.752	0.817
Dependency Tree-LSTM (Tai et al., 2015)	0.769	0.824
Constituent Tree-LSTM (Tai et al., 2015)	0.760	0.803
Siamese LSTM (Mueller et al., 2016)	0.767	0.820
InferSent (Conneau et al., 2017)	0.762	0.831
Avg. BERT embeddings (Devlin et al., 2019)	0.718	0.737
Text-GNN (Huang et al., 2019)	0.776	0.835
GCSTS	**0.785**	**0.854**

Table 3 An ablation study for the proposed model.

GCSTS	STS15(p)	MRPC(Acc.)
(1)- Iterated GCN blocks	0.723	0.765
(2)- Dynamic graph structures	0.730	0.769
(3)- BiLSTM encoding layer	0.720	0.774
(4)- POS embedding	0.742	0.784

Results (continued)

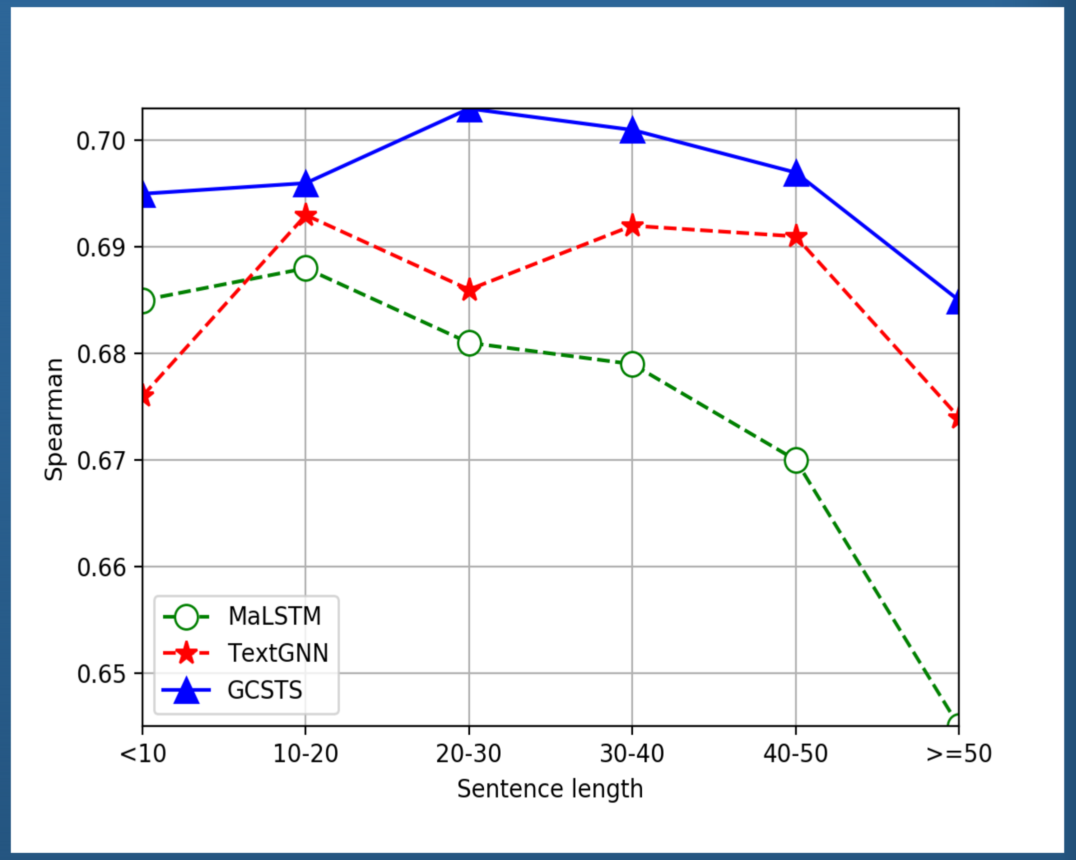

Figure 2 Comparison of GCSTS, MaLSTM, and Text-GNN against different sentence lengths.

We use the same metrics as previous work [9], [17-23] to evaluate the performance of the proposed model. We report the Spearman correlation coefficient (p) for short text similarity on STS(2012-2016) and STS-B datasets. The accuracy and F1 score are used on the MRPC dataset. Table I and Table II report the results of our models against other baseline methods. We can see that our model can achieve the best results on 6/7 datasets. As is shown in the table, the results of graph-based models are better than traditional models like CNN, LSTM. Compared with these sequence- based models, graph-based models extract not only semantic information but also structure information of sentences. Our model also performs better than graph-based models like Tree-LSTM and Text-GNN. Tree-LSTM based models are usually difficult to parallelize and thus computationally inefficient.

Conclusion

- In this paper, we propose a GCNs-based context-aware short text similarity model. GCSTS uses iterated GCN blocks and non-fixed neighbor matrices to incorporate more contextual information of sentences. The experimental results show that the proposed model can effectively improve the performance of sentence similarity measurement and provides a new way to train deeper GCNs.

Acknowledgement

- This work is supported by the State Key Program of National Nature Science Foundation of China (No. 61936001) and the National Natural Science Foundation of China (No.52073169). We also appreciate the High Performance Computing Center of Shanghai University, and Shanghai Engineering Research Center of Intelligent Computing System (No. 19DZ2252600) for providing the computing resources and technical support.